METHOD FOR ACHIEVING HIGHER AVAILABILITY OF COMPUTER PCI ADAPTERS

Background of the Invention

5   1.   Field of the Invention

The invention relates to a method for achieving higher availability of

computer PCI adapters, and specifically for achieving higher availability by

dynamically switching PCI adapters between host-controlled operation and IOP-

controlled operation.

10   2.   Background Information

A PCI (Peripheral Component Interconnect) adapter allows a computer to

interface using a standard protocol with a variety of peripheral devices to input and

output data to the computer. In current computer architectures, PCI adapters operate

in only one of two modes, called "native" and "non-native." A PCI adapter that

15   operates under the control of the system processor is termed host-controlled, or

"native". A PCI adapter that operates under the control of a separate IOP (Input-

Output Processor) is termed IOP-controlled, or "non-native." For example, in the

case of a non-native PCI adaptor, it becomes necessary to turn off all of the processes

that are running on that PCI adapter, if it is desired to service the IOP. This may be

20   for a short time if the replacement IOP is available and the concurrent maintenance

action is performed right away. In other cases, the system could be down for a

significant period of time awaiting replacement of the IOP.

Another problem inherent in current PCI adapters, which are able to operate

only in either the "native" or "non-native" mode, is the lack of any capability for

balancing the work load between the system processor and an IOP. If, for example, the PCI adaptor is running natively, i.e., under the system processor, and an I/O intensive job causes the overall system performance to decline, there is no current capability for switching the PCI adapters to run non-natively so as to off-load some of

5   the work from the system processor to an IOP. Similarly, there is also no way in present systems to dynamically move non-native PCI adapters from an IOP that is experiencing high utilization so that they can run natively.

Summary of the Invention

It is, therefore, a principle object of this invention to provide a method for

10   achieving higher availability of computer PCI adapters.

It is another object of the invention to provide a method for achieving higher availability of computer PCI adapters that solves the above-mentioned problems.

These and other objects of the present invention are accomplished by the method for achieving higher availability of computer PCI adapters disclosed herein.

15   In an exemplary aspect of the invention, a computer system includes a system processor, an input/output processor, and an input/output adapter connected to the system processor and the input/output processor. In order to achieve higher availability of the computer input/output adapter, the input/output adapter is capable of dynamically switching between being controlled by the system processor ("native"

20   mode) and being controlled by the input/output processor ("non-native" mode). The input/output adapter is preferably a PCI (Peripheral Component Interconnect) adapter, and the input/output processor is preferably a PCI-compatible processor. The

computer system may include more than input/output adapter that is dynamically switchable.

In another aspect of the invention, a method for fault recovery in a computer system having a system processor, an input/output processor, and the above-described

5   dynamically switchable input/output adaptor, includes the steps of: detecting a fault in the input/output processor; and switching the input/output adapter to control by the system processor if the input/output adapter is being controlled by the input/output processor when the fault is detected. The fault recovery method may also include the further steps of: detecting correction of the fault in the input/output processor; and

10   switching the input/output adapter back to control by the input/output processor when the correction of the default is detected, if it was previously switched to control by the system processor as a result of the original fault in the input/output processor. The method is also applicable to a computer system have a plurality of dynamically switchable input/output adapters, in which case each of the dynamically switchable

15   input/output adapters being controlled by the input/output processor is switched to control by the system processor when the fault is detected, and each of the dynamically switchable input/output adapters being controlled by the system processor when correction of the fault is detected is switched back to control by the input/output processor if it was previously switched to control by the system

20   processor as a result of the original fault in the input/output processor.

In a further aspect of the invention, a method for optimizing processor utilization in a computer system having a system processor, an input/output processor, and a dynamically switchable input/output adaptor, includes the steps of: determining computer system utilization using a commonly-known measurement tool; and

switching control of the input/output adapter from the then controlling processor to

the non-controlling processor, if it is determined that the controlling processor is

being over utilized and the non-controlling processors has sufficient capacity that

switching control of the input/output adapter will not adversely affect overall system

5   throughput. Preferably, switching control of the input/output adapter from the

controlling processor to the second one of the processor is also based on a

determination that the over utilization of the controlling processor is likely to continue

for at least a specified period of time. The steps of determining computer system

utilization and switching control of the input/output adapter based on such

10   determination are preferably repeated at intervals substantially equal to the specified

period of time. The method is also applicable to a computer system have a plurality

of dynamically switchable input/output adapters, in which case the steps of

determining computer system utilization and switching control of the input/output

adapter based on such determination are performed for each of the plurality of

15   input/output adapters.

        Brief Description of the Drawings

        Figure 1 illustrates in simplified form the architecture of a typical computer

system.

        Figure 2A show schematically the switchable dual-mode input/output adapters

20   of the present invention operating under control of an input/output processor.

        Figure 2B show schematically the switchable dual-mode input/output adapters

of the present invention operating under control of the system processor.

Figure 3 is a flow chart illustrating the steps to be performed in an IOP maintenance procedure according to the present invention.

Figure 4 is a flow chart illustrating the steps to be performed in load balancing the system and input/output processors according to the present invention.

Detailed Description of the Preferred Embodiments

The invention will now be described in more detail by way of example with reference to the embodiments shown in the accompanying figures. It should be kept in mind that the following described embodiments are only presented by way of
5   example and should not be construed as limiting the inventive concept to any particular physical configuration.

Further, if used and unless otherwise stated, the terms "upper," "lower," "front," "back," "over," "under," and similar such terms are not to be construed as limiting the invention to a particular orientation. Instead, these terms are used only on
10   a relative basis.

The present invention is directed toward a method of achieving higher availability of PCI adapters by dynamically switching the PCI adapters between host-controlled operation and IOP-controlled operation.

Figure 1 illustrates in simplified form the architecture of a typical computer
15   system 10. The typical computer system 10 includes: a system processor 1 for running system routines, such as input/output functions, and applications programs; a memory 2 for storing the system and applications programs and data used by the processor 2; and one or more input/output adapters (IOAs) 3a-3c for interfacing with input/output and storage devices, such as hard disk drives, floppy disk and CD-ROM

readers, and printers. The IOAs may, for example, be PCI (Peripheral Component

Interconnect) adapters, each of which may be adapted to exchange data with a

different kind of device, but which all operate using a standard input/output protocol.

The elements of the computer system 10 are typically connected with a bus 5,

5   over which instructions and data flow between them. Although only one bus 5 is

shown for simplicity, a computer system may include a number of separate buses for

input/output, memory and other specialized functions. PCI adapters are typically

single printed-circuit board devices with standard physical attributes so that they can

be readily plugged into the computer system 10 to interface with a PCI-compatible

10   bus.

The typical computer system 10 may also include an input/output (I/O)

processor (IOP) 4, which is intended to perform the specialized function of

controlling the IOAs, thus relieving the system processor of that function. The IOP

may, for example, be a single printed-circuit card processor compatible with the PCI

15   standard that plugs into the computer system 10 to interface with the same bus as the

IOAs.

To overcome the problems inherent in the types of PCI adapters currently in

use, which operate only in either the host-controlled ("native") mode or the IOP-

controlled ("non-native") mode, the present invention makes use of PCI adapters, or

20   more generally IOAs, capable of dynamically switching between the native and non-

native modes. Figure 2A shows schematically a configuration in which the IOP 4 is

controlling a number of such dual-mode IOAs 13a-13c that can switch between being

controlled by an IOP non-natively and being controlled by the system natively. If the

IOP experiences problems, then, without the present invention, all the IOAs assigned

to that IOP would become nonfunctional. With the present invention, the IOAs 13a-13c can be switched to run in native mode so as to allow the IOAs 13a-13c to continue functioning, as depicted schematically in Figure 2B. The faulty IOP 4 can then be replaced and, if desired, the IOAs 13a-13c can be reassigned to the IOP 4. A

5   fault recovery scheme can also be implemented in which the IOAs 13a-13c are switched automatically in the event a failure of IOP 4 is detected.

Figure 3 is a flow chart illustrating the steps to be performed in the IOP maintenance procedure briefly discussed above. The computer is preferably one in which concurrent maintenance, or "hot swapping," of PCI cards can be performed

10   without turning off system power. After starting the maintenance procedure at step S1, a determination is made at step S2 as to whether the IOP in question controls switchable, dual-mode IOAs. If the answer is yes, then at step S3 the switchable IOAs are switched to system control in order to allow the input/output functions they have been performing to continue.

15   Maintenance on the IOP in question is performed at step S4, after which a determination is made at step S5 as to whether any IOAs were switched for the maintenance action. If so, then at step S6 the switched IOAs are switched back to IOP control. At step S7 the maintenance action is complete.

A variation of the foregoing method provides for automatic fault recovery in

20   the event of an IOP failure or malfunction. This requires that in step S1 the system be capable of detecting such fault to initiate the maintenance procedure, and in step S3, automatically switching over to system control the IOAs controlled by the faulty IOP. The system would also have to be capable in step S4 of detecting when the faulty IOP had been replaced, and hence was once again operational, and in step S6,

automatically initiating the process of switching back the IOAs that had originally been switched because of the fault.

Another aspect of the present invention is the capability of automatically switching between native and non-native modes based on the level of demand or

5   utilization of the processors in the system. Figure 4 is a flow chart illustrating the steps to be performed in such an optimization process. After initiating the process in step S10, an existing performance tool, one known to those of ordinary skill in the art, would be used in step S11 to determine system utilization. If, at step S12, it is determined that the processor (either the system processor or the IOP) currently

10   controlling an IOA is being over utilized and it can be determined that this over utilization is likely to continue for a period of time, then the IOA is a candidate for switching. To be switched, an affirmative determination must also be made at step S13 that the non-controlling processor to which the candidate IOA would be switched, has enough capacity that the switch would not adversely affect the current

15   throughput of the system. For example, if an IOA is operating in native mode and is processing a job that is input/output intensive and could take a significant amount of time to complete, then it would be beneficial to reassign this IOA to an IOP that is not being heavily utilized and allow the system processor to do other work.

Affirmative determinations in both steps S12 and S13 will result in switching

20   the candidate IOA to the non-controlling processor in step S14. Thereafter, the system waits an appropriate period of time in step 15, which may be substantially the same as the time during which over utilization of the previously controlling processor is likely to continue, as specified in step S12. The process then starts again at step S11.

A variation of the foregoing process involves consideration of each of the dual-mode IOAs as a candidate for switching if the system includes more than one such dual-mode IOA.

5   It should be understood that the invention is not necessarily limited to the specific process, arrangement, materials and components shown and described above, but may be susceptible to numerous variations within the scope of the invention. For example, although the above-described exemplary aspects of the invention are believed to be well suited for general-purpose computers, it is contemplated that the

10   concepts of the present invention can be applied in other types of computer systems, such as servers and communications processors. In general, the concepts of the present application can be utilized whenever it is desired to optimize data input and output in a processor-based system.

It will be apparent to one skilled in the art that the manner of making and

15   using the claimed invention has been adequately disclosed in the above-written description of the preferred embodiments taken together with the drawings.

It will be understood that the above description of the preferred embodiments of the present invention are susceptible to various modifications, changes and adaptations, and the same are intended to be comprehended within the meaning and

20   range of equivalents of the appended claims.